INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY
THIRUVANANTHAPURAM

Assignment #4

Due on 05-11-2014

SUHAS S
(SC14M081)

Contents

1.Ridge Regression

1.1 2nd degree polynomial fitting

λ v/s training & validation error.

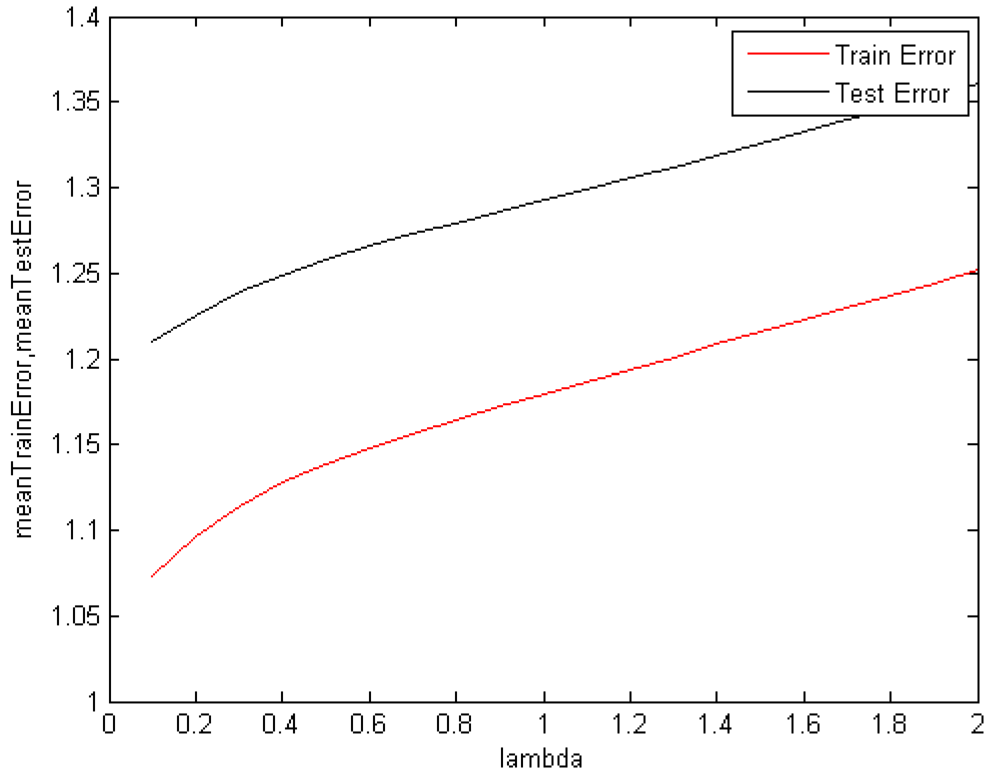

Figure 1: fitting 2nd degree polynomial

Plot of J_reg(W)

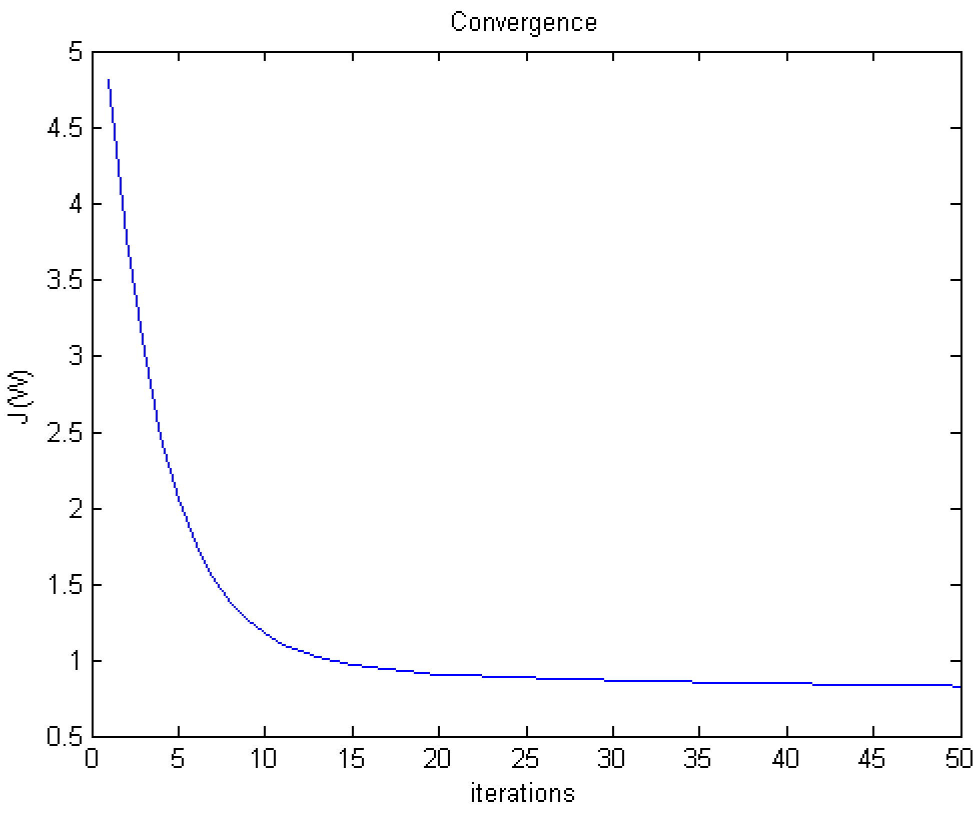

Figure 2: J(W) v/s iterations

1.2 3rd degree polynomial fitting

λ v/s training & validation error.

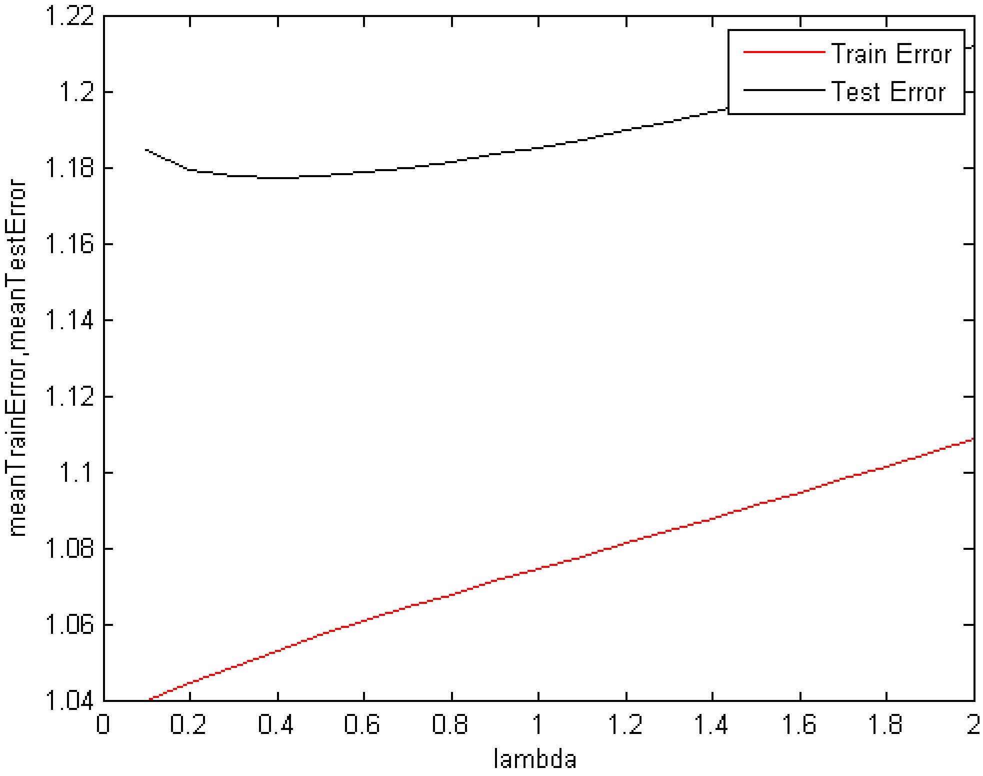

Figure 3: fitting 3rd degree polynomial

Plot of J_reg(W)

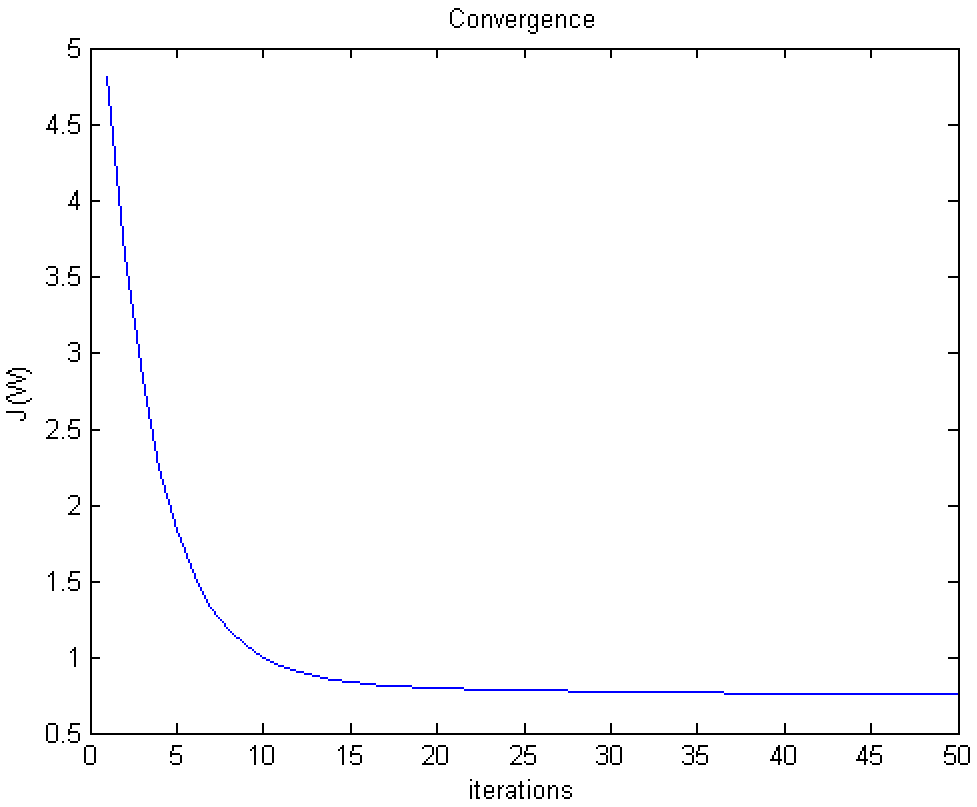

Figure 4: J(W) v/s iterations

1.3 7th degree polynomial fitting

λ v/s training & validation error.

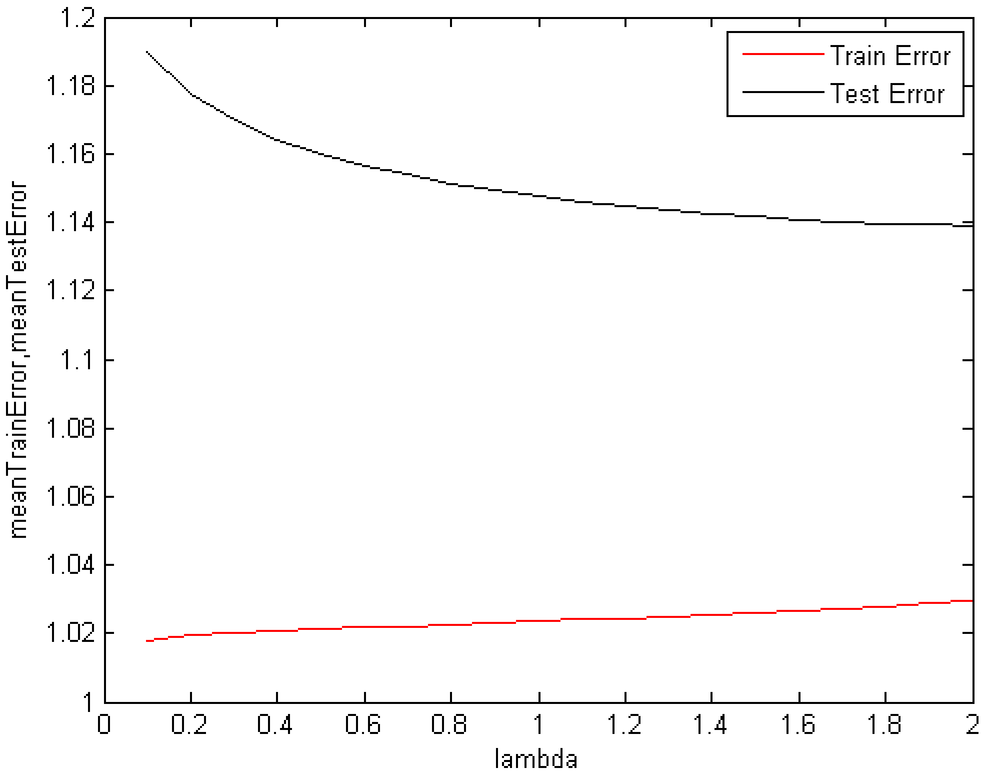

Figure 5: fitting 7th degree polynomial

Plot of J_reg(W)

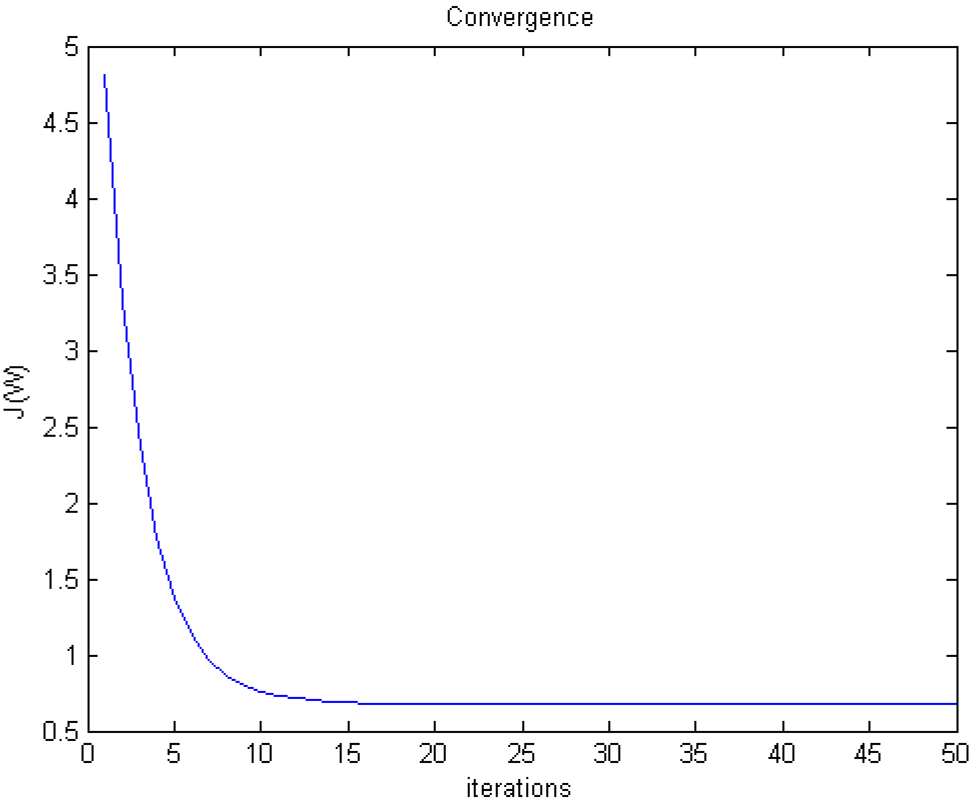

Figure 6: J(W) v/s iterations

1.4 Performance Comparisons

best model with 2nd degree polynomial : λ=0.10 test error estimated=1.2096
best model with 3rd degree polynomial : λ=0.40 test error estimated=1.1777
best model with 7th degree polynomial : λ=2.00 test error estimated=1.1389
performance of the least square method : 1.4270
Thus 7th degree polynomial gives the best fit for the data.

2.Regularised linear regression

Weight values(without regularisation) θ_0 = 11.6506 θ_1 = -1.7925 θ_2 = 4.4062 θ_3 = -1.6779 θ_4 = 4.1600 θ_5 = -0.5733 θ_6 = 16.4432 θ_7 = 1.6647 θ_8 = 0.3262 θ_9 = 0.9794 θ_{10} = -2.1768 θ_{11} = -4.6693 θ_{12} = 8.5775 θ_{13} = -10.4561

Weight values(with regularisation) θ_0 = 11.9142 θ_1 = -1.8099 θ_2 = 4.4106 θ_3 = -1.6696 θ_4 = 4.1478 θ_5 = -0.5509 θ_6 = 15.9588 θ_7 = 1.7438 θ_8 = 0.3979 θ_9 = 0.9771 θ_{10} = -2.1800 θ_{11} = -4.6242 θ_{12} = 8.5004 θ_{13} = -10.6712

cost witout regularisation: 8.37
cost with regularisation: 8.20
α value : 0.500000
λ value : 0.10

3.K-nearest neighbourhood

no of test datasets: 209
no of folds: 5
no of misclassifications: 8
accuracy: 0.962
precision: 0.944
recall/sensitivity: 0.944
F-Measure: 0.944
Max accuracy during cross validation: 0.956
optimum K value: 1

The ROC-curve obtained is shown below

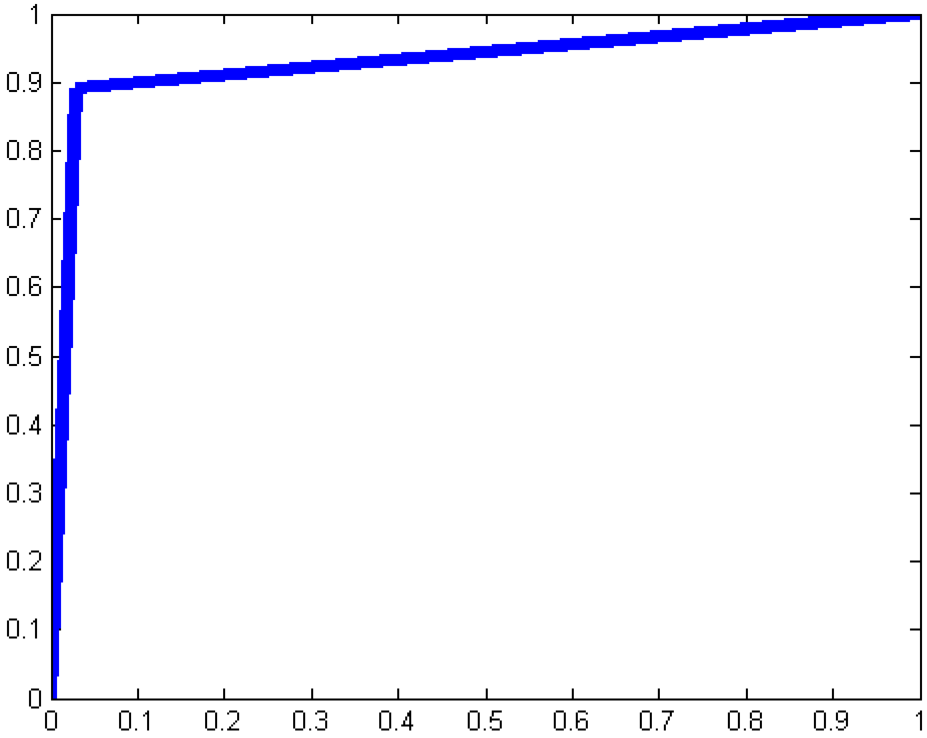

Figure 7: ROC curve for k=1

4.Decision Tree in WEKA

=== Evaluation on test split ===

Correctly Classified Instances: 5584 (85.7494%)
Incorrectly Classified Instances: 928 (14.2506%)
Kappa statistic: 0.5732

```
K&B Relative Info Score: 288671.6144
K&B Information Score: 2308.5563 bits              0.3545 bits/instance
Class complexity | order 0 :  5100.5935 bits            0.7833 bits/instance
Class complexity | scheme :  40494.5293 bits              6.2184 bits/instance
Complexity improvement(Sf): -35393.9358 bits                -5.4352 bits/instance
Mean absolute error: 0.1917
Root mean squared error: 0.3191
Relative absolute error: 52.867%
Root relative squared error: 75.4903%
Total Number of Instances: 6512
```

=== Detailed Accuracy By Class ===

TP Rate	**FP Rate**	**Precision**	**Recall**	**FMeasure**	**ROC Area**	**Class**
0.936	0.4	0.885	0.936	0.91	0.89	<=50K
0.6	0.064	0.739	0.6	0.662	0.89	>50K
Weighted Avg						
0.857	0.322	0.851	0.857	0.852	0.89	-

=== Confusion Matrix ===

Class<=50k	**Class>50k**
4674	322
606	910

The Complete set of results with the obtained decision tree is accessible in this link(since it is around 1000 lines it is not included here)

5.Problem on Apriori Algorithm

Support counts of individuals

M	O	N	K	E	Y	D	A	U	C	I
3	4	2	5	4	3	1	1	1	2	1

Since Min.Support is 3(60%) we form L1 as

M	O	K	E	Y
3	4	5	4	3

Then C2 is formed as shown below

M,O	M,K	M,E	M,Y	O,K	O,E	O,Y	K,E	K,Y	E,Y
1	3	2	2	4	4	2	4	3	2

L2 obtained is

M,K	O,K	O,E	K,E	K,Y
3	4	4	4	3

Then C3 is formed as shown below

M,K,E
4

C3 is same as L3 since its support count is 4(>=3).
The rules formed are

$\{O,K\} \implies E$
$\{O,E\} \implies K$
$\{K,E\} \implies O$

$\frac{O,K,E}{O,K} = \frac{4}{4} = 1$
$\frac{O,K,E}{O,E} = \frac{4}{4} = 1$
$\frac{O,K,E}{K,E} = \frac{4}{4} = 1$

since confidence >=80%, all are strong associations

Output of WEKA tool

Generated sets of large itemsets:
Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 1

Best rules found:

- E=YES 4 ==> K=YES 4 conf: (1)
- D=NO 4 ==> K=YES 4 conf: (1)
- A=NO 4 ==> K=YES 4 conf: (1)
- U=NO 4 ==> K=YES 4 conf: (1)
- I=NO 4 ==> K=YES 4 conf: (1)
- U=NO 4 ==> E=YES 4 conf: (1)
- E=YES 4 ==> U=NO 4 conf: (1)
- E=YES U=NO 4 ==> K=YES 4 conf: (1)
- K=YES U=NO 4 ==> E=YES 4 conf: (1)
- K=YES E=YES 4 ==> U=NO 4 conf: (1)